The Korean Communications
in Statistics Vol. 10 No. 3, 2003
pp. 1069-1086

Multistage Point and Confidence Interval Estimation of the Shape Parameter of Pareto Distribution

H. I. Hamdy[1] M. S. Son[2] M. K. Gharraph A. M. Rashad

Abstract

This article presents the asymptotic theory of triple sampling procedure as pertain to estimating the shape parameter of Pareto distribution. Both point and confidence interval estimation are considered within the same inference unified framework. We show that this group sampling technique possesses the efficiency of Anscome (1953), Chow and Robbins (1965) purely sequential procedure as well as reduce the number of sampling operations by utilizing Stein (1945) two stages procedure. The analysis reveals that the technique performs excellent as far as the accuracy is concerned. The present problem differs from those considered by many authors, in multistage sampling, in that the final stage sample size and the parameter's estimate become highly correlated and therefore we adopted different approach.

Keywords. Cost of perfect information; estimation cost; foregone risk; sampling cost; triple sampling.

1. Introduction

Pareto distribution plays an important role in so many areas of application, including investigations of city population sizes, stock price fluctuations, insurance risk, error clustering in communication circuits, natural resources, business failure, life testing and reliability analysis (see for example Mann, Schaffer and Singpurawalla (1974, pp 146)). In statistical quality control, Pareto distribution is used to study the cause and effect relation between different types of failures and is known as Pareto chart. Pareto distributions also arise as limiting distributions to problems involve waiting times and life testing (see for example Ahsanullah and Houchens (1989), Ahsanullah (1991), and Balakrishnan and Ahsanullah (1994)). The model was first introduced by Pareto (1897) to study the distribution of income. The distribution is also related to Lorenz curve and Gini index. Bayesian statistical inference procedures for Pareto parameters were considered by Muniruzzaman (1963), Malik (1970), Sinha and Howlader (1980), Arnold and Press (1983), Nigm and Hamdy (1987), Arnold and Presss (1989). Liang (1993) considered the case of estimating the scale parameter σ, when the shape parameter θ is known under the squared error loss function using nonparametric Bayesian statistical inference. For a detailed list of reference on parameters estimation, characterization, and related distributions we refer to Johnson, Kotz and Balakrishnab (1994, pp 620-627).

[1] Department of Quantitative Methods and Information Systems, KU, P.O.Box 5486, Safat 13005 Kuwait
[2] Department of Mathematics and Statistics, 16 Colchester Avenue, UVM, Burlington, VT 05401, USA
E-mail : son@emba.uvm.edu

To address the problem let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed random variables having the first kind Pareto distribution of the form

$$f(x;\theta,\sigma) = \sigma^{-1}\theta^{\sigma^{-1}} x^{-\sigma^{-1}}, \quad x \geq \theta > 0, \sigma > 0. \tag{1.1}$$

The scale parameter σ and the shape parameter θ are assumed unknown. Having recorded a random sample of size n from the distribution in (1.1), we let

$$X_{n(1)} = \min_{1\leq i\leq n} X_i, \text{ and } \hat{\sigma}_n = n^{-1}\sum_{i=1}^{n} ln\left(X_i / X_{n(1)}\right)$$

be the MLE's point estimators for θ and σ respectively. However, we will use the unbiased version $\hat{\sigma}_n = (n-1)^{-1}\sum_{i=1}^{n} ln\left(X_i / X_{n(1)}\right)$ to develop theory in subsequent sections.

In this study we focus on estimating the shape parameter σ under the following quadratic loss function with linear sampling cost

$$L_n(A) = A(\hat{\sigma}_n - \sigma)^2 + cn \tag{1.2}$$

where $A > 0$ is a weight constant to be determined later in this section, and c is the cost per unit sample which assumed to be known. The first term in the right hand side of (1.2) reflects the cost due to estimation, while the second term is known as the sampling cost. The risk $R_n(A)$ associated with (1.2) is then obtained by taking the expectation of both sides as

$$E(L_n(A)) = AE(\hat{\sigma}_n - \sigma)^2 + cn$$

$$R_n(A) = A\sigma^2 n^{-1} + cn \tag{1.3}$$

Treating n in (1.3) as a continuous variable, we differentiate (1.3) with respect to n and obtain the optimal sample size that minimizes the encountered risk as

$$n^* = \sqrt{A/c}\ \sigma \tag{1.4}$$

Assume further that a fixed $2d$, say; width confidence interval for σ is required such that the coverage probability that σ lies within the interval is at least the nominal value $100(1-\alpha)\%$. Since $\hat{\sigma}_n$ is the MLE for σ, it follows that $\dfrac{(\hat{\sigma}_n - \sigma)\sqrt{n}}{\sigma} \xrightarrow{\ell} N(0, 1)$, for sufficiently large n. Therefore under the normality assumption and the requirement that $P(|\hat{\sigma}_n - \sigma| \leq d) \geq 1-\alpha$, the optimal sample size required to construct the interval should be

$$n \geq n^* = (a^2/d^2)\,\sigma^2 \tag{1.5}$$

where a is the upper $100(1-\alpha/2)\%$ cutoff point of the standard normal distribution. The quantity (a^2/d^2) appears in the right hand side of (1.5) is Fisher's optimal information, which represents the amount of information required to explore a unit variance in order to construct a $2d$ fixed width confidence interval for σ with at least $100(1-\alpha)\%$. Hence, comparing n^* in (1.4) and (1.5) we have, $A=(a^2/d^2)\,cn^*$ which refers to the cost of perfect information. While cn^* refers to the optimal cost of sampling. By the cost of perfect information we mean, the amount of information required to explore a unite variance in order to attain the optimal risk.

Since σ in (1.4) is unknown n^* is also unknown. On facing the ignorance of σ in (1.4), we resort to the following three-stage method of sampling to estimate σ via estimation of n^*. The existing literatures in multistage sampling (purely sequential, two stage, three-stage and accelerated sequential schemes) are limited and focused primarily on estimating the shape parameter θ (see for example Wang (1973), Mukhopadhyay and Eko (1985), and Hamdy and Pallotta (1987) for details). However, we are not aware of any publications in multistage inference about σ. The reason behind such limitation is rather technical.

In section 2 we describe the triple sampling procedure and its asymptotic characteristics along the lines of Hall (1981), Mukhopadhyay (1985), Mukhopadhyay et al. (1987), Hamdy and Pallotta (1987), Hamdy (1988), Hamdy et al. (1988) and Hamdy et al. (1989). We also stress that the techniques adopted by the above authors are different from those presented in this paper, for the dependency between the three-stage stopping rule and the estimate of σ becomes troublesome. Therefore the analysis is different from those authors listed above. Nevertheless, the decision frame in the present work is similar to Almahmeed et. al. (2002).

2. Triple Sampling Estimation Procedure for σ

Sequential sampling in three-stages was first introduced in the seminal work of Hall (1981) to construct fixed width confidence interval for the normal mean with a predetermined confidence coefficient $100(1-\alpha)\%$. His procedure is designed to combine the efficiency of the purely sequential procedures and the operational savings made possible by sampling in batches. The originality of his procedure and the elegance of the sampling technique are astonishing.

As the name might suggest, the three-stage estimation procedure takes three sampling phases to finish the whole estimation process namely, the pilot study phase, the main study phase and the fine tuning phase. The pilot study phase, starts by taking an initial pilot sample of size $m \geq 2$, say, from the distribution in (1.1) to initiate an estimate for the unknown parameter σ. Denote these potential observations by $X_1, X_2, \ldots, X_m$. Then we compute $\hat{\sigma}_m = \sum_{i=1}^{m} ln\left(X_i / X_{m(1)}\right)$, where $X_{m(1)} = \min_{1\leq i\leq m} X_i$. In the main study phase, we estimate, only, a fraction $\gamma \varepsilon (0,1)$ of n^*. Then we mimic the form of n^* given in (1.4) and define the following second stage sample size as

$$N_1 = max\left\{ m, \left[\gamma a^2 \hat{\sigma}_m^2 / d^2 \right] + 1 \right\} \tag{2.1}$$

where $[x]$ is the largest integer less than or equal to x. If $m > [\gamma a^2 \hat{\sigma}_m^2 / d^2] + 1$, we stop at this stage. Otherwise, we continue to observe an additional sample of size $N_1 - m$ from the distribution in (1.1) and update the estimate of σ based on the augmented N_1 observations. We then, define the final stage sample size as

$$N = max\left\{ N_1, \left[a^2 \hat{\sigma}_{N_1}^2 / d^2 \right] + 1 \right\} \tag{2.2}$$

We terminate the whole sampling process at this stage, if $N_1 > \left[a^2 \hat{\sigma}_{N_1}^2 / d^2 \right] + 1$. Differently, we proceed to record $N - N_1$ extra samples to be augmented with the previously selected N_1 samples. Upon terminating the sampling procedure we construct the confidence interval $(\hat{\sigma}_N - d, \hat{\sigma}_N + d)$ and propose $\hat{\sigma}_N$ as the point estimate for the unknown shape parameter σ to minimize the anticipated risk. It is clear from the final stage, rule in (2.2), that the final sample size N and $\hat{\sigma}_N$ are highly correlated and therefore the analysis takes somewhat different turn. The following Lemma 1 gives preliminary results concerning the three-stage procedure (1.2)-(2.2).

Lemma 1 *In the context of the three-stages rule (2.1)-(2.2) we have as* $d \to 0$,

$(i) E(\hat{\sigma}_{N_1}) = \sigma - d^2(\gamma a^2)^{-1} + o(d^2)$

$(ii) Var(\hat{\sigma}_{N_1}) = \sigma d^2(\gamma a^2)^{-1} + o(d^2)$

$(iii) E|\hat{\sigma}_{N_1} - \sigma|^3 = o(d^2)$.

$(iv) E\, h(\hat{\sigma}_{N_1}) = h(\sigma) + \frac{1}{2} d^2 a^{-2} \gamma^{-1} + o(d^2)$, *for some continuously differentiable bounded function* $h(\cdot)$.

$(v) Var(h(\hat{\sigma}_{N_1})) = \gamma^{-1} d^2 a^{-2} + o(d^2)$.

The following transformation enables one to replace the random variables N_1 and N by other random variables M_1 and M respectively, which are defined similarly but in terms of the sample mean of *i.i.d.* random variables. Let $U_i = ln(X_i)$ for $i = 1, 2, 3, \ldots$ It follows that the random variables $U_1, U_2, U_3, \ldots$ are *i.i.d.* random variables with the following probability density function

$$k(u;ln(\theta),\sigma)=\sigma^{-1}e^{-(u-ln(\theta))/\sigma} \qquad u>ln(\theta),\ \sigma>0. \tag{2.3}$$

It follows that

$$\hat{\sigma}_n=\sum_{i=2}^{n}(n-i+1)(U_{(i)}-U_{(i-1)})/(n-1),\quad n=2,3,...$$

Where $U_{(i)}$ is the i^{th} order statistic from (2.3) and $U_{(0)}=0$. Finally, we implement the following common transformation technique $Z_i=(n-i+1)(U_{(i)}-U_{(i-1)})$. It follows that the random variables Z_i are *i.i.d.* with the following probability density function

$$k(z;\sigma)=\sigma^{-1}e^{-z/\sigma} \qquad z>0,\ \sigma>0.$$

Hence, we write $\hat{\sigma}_n$ as $\bar{Z}_n=\sum_{i=1}^{n-1}Z_i/(n-1),\ n=2,3,...$

It follows from Lemma (6.1) of Lombarod and Swanepoel (1977) (see also Swanepoel and Van Wyke (1982)) that decision rules based on the sequences $\{\hat{\sigma}_n^2,\ n\geq 2\}$ and $\{\bar{Z}_n^2,\ n\geq 2\}$ are equivalent. Consequently, we define

$$M_1=max\left\{m\ ,\ \left[\gamma a^2\bar{Z}_m^2/d^2\right]+1\right\} \tag{2.4}$$

$$M=max\left\{M_1\ ,\ \left[a^2\bar{Z}_{M_1}^2/d^2\right]+1\right\} \tag{2.5}$$

It follows that the random variables N_1 and M_1 as well as N and M are identically distributed. Therefore, we continue to use the stopping rules (2.4)−(2.5) instead of (2.1)−(2.3) to develop our subsequent results.

Proof of Lemma 1 To prove (i), we write $E(\hat{\sigma}_{N_1})=E(\bar{Z}_{M_1})$. Then conditioning on the $\sigma-field$ generated by the random variables $Z_1,Z_2,...,Z_{m-1}$ to have

$$\begin{aligned}E(\bar{Z}_{M_1})&=EE(\bar{Z}_{M_1}|Z_1,...,Z_{m-1})\\&=E\left\{(M_1-1)^{-1}E(\sum_{i=1}^{M_1-1}Z_i|Z_1,...,Z_{m-1})\right\}\\&=E\left\{(M_1-1)^{-1}E((\sum_{i=1}^{m-1}Z_i+\sum_{i=m}^{M_1-1}Z_i)\,|\,Z_1,...,Z_{m-1})\right\}\end{aligned}$$

Given $Z_1, Z_2, \ldots, Z_{m-1}$, the first sum $\sum_{i=1}^{m-1} Z_i$ is non-random, hence,

$$\begin{aligned} E(\bar{Z}_{M_1}) &= E\left\{(M_1-1)^{-1}\left(\sum_{i=1}^{m-1} Z_i + \sigma(M_1-m)\right)\right\} \\ &= \sigma + E\left\{(m-1)(M_1-1)^{-1}(\bar{Z}_m - \sigma)\right\} \end{aligned} \tag{2.6}$$

Expanding $(M_1-1)^{-1}$ around γn^* using Taylor series,

$$(M_1-1)^{-1} = (\gamma n^*)^{-1} - (M_1 - \gamma n^*)(\gamma n^*)^{-2} + (M_1 - \gamma n^*)^2 \zeta^{-3} \tag{2.7}$$

where ζ appears in (2.7) is a random variable lies between γn^* and M_1. For large d, we approximate

$$M_1 \approx \gamma a^2 d^{-2} \bar{Z}_m \tag{2.8}$$

To see that, we write $M_1 = \left[\gamma a^2 d^{-2} \bar{Z}_m\right] + 1$, *a.s.* as $d \to 0$ except possibly on some set $\xi = \left(\gamma a^2 d^{-2} \bar{Z}_m \le m\right)$, of measure zero, where $h = h(d) = 2m(m-1)d^2 / \gamma a^2$. Clearly as $d \to 0$, $g \to 0$. Hence,

$$\begin{aligned} P\left(\gamma a^2 d^{-2} \bar{Z}_m \le m\right) &= P\left(2(m-1)\bar{Z}_m / \sigma \le h\right) \\ &= P(\chi^2_{2(m-1)} \le h) \\ &= \int_0^h t^{m-2} e^{-\frac{t}{2}} dt / \Gamma(n-1) \\ &\le \int_0^h t^{m-2}\, dt / \Gamma(m) \\ &= h^{m-1} / \Gamma(m) \\ &= o(1) \end{aligned}$$

Thus, we have

$$M_1 = \left(\left[\gamma a^2 d^{-2} \bar{Z}_m\right] + 1\right) I(\gamma a^2 d^{-2} \bar{Z}_m > m) + m\, I(\gamma a^2 d^{-2} \bar{Z}_m \le m)$$

where $I(\cdot)$ is the set indicator function. Therefore, we write

$$M_1 = \left(\gamma a^2 d^{-2} \bar{Z}_m + 1 - (\gamma a^2 d^{-2} \bar{Z}_m - \left[\gamma a^2 d^{-2} \bar{Z}_m\right]\right) I(\gamma a^2 d^{-2} \bar{Z}_m > m) + m\, I(\gamma a^2 d^{-2} \bar{Z}_m \le m)$$

Let M_{1C} be the continuous version of M_1, i.e. $M_{1C} = max\left(m, \left(\gamma a^2 d^{-2} \overline{Z}_m\right)\right)$, then we write

$$M_{1C} = \gamma a^2 d^{-2} \overline{Z}_m \, I(\gamma a^2 d^{-2} \overline{Z}_m > m) + m I(\gamma a^2 d^{-2} \overline{Z}_m \le m)$$

It follows that

$$M_1 - M_{1C} = \beta I(\gamma a^2 d^{-2} \overline{Z}_m > m)$$

where $\beta = 1 - \left(\gamma a^2 d^{-2} \overline{Z}_m - \left\lfloor \gamma a^2 d^{-2} \overline{Z}_m \right\rfloor\right)$ is uniformly distributed over $(0,1)$, (see for example Hall (1981), for details).

Now,

$$\begin{aligned} E(M_1 - M_{1C}) &= E\left\{\beta \, I(\gamma a^2 d^{-2} \overline{Z}_m > m)\right\} \\ &\le \left(E\left(\beta^2\right) E^2 \, I\left(\gamma a^2 d^{-2} \overline{Z}_m > m\right)\right)^{\frac{1}{2}} \\ &= \left(E\left(\beta^2\right) P^2 \left(\gamma a^2 d^{-2} \overline{Z}_m > m\right)\right)^{\frac{1}{2}} , \\ &\le \left(E\left(\beta^2\right)\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{3}\right)^{\frac{1}{2}} = 0.58 \end{aligned}$$

where we have used Cauchy–Schwartz inequality.
As indicated, even for moderate values of d, the expected difference between M_1 and M_{1C} is less than 0.58 of an observation. It follows that

$$(M_1 - 1)^{-1} = (\gamma n^*)^{-1} - \gamma a^2 d^{-2} (\overline{Z}_m - \sigma)(\gamma n^*)^{-2} + (\gamma a^2)^2 d^{-4} (\overline{Z}_m - \sigma)^2 \zeta^{-3}.$$

Thus, (2.6) yields

$$\begin{aligned} E\left(\overline{Z} M_1\right) &= \sigma + E\left\{(m-1)(\overline{Z}_m - \sigma)\left\{(\gamma n^*)^{-1} - \gamma a^2 d^{-2} (\overline{Z}_m - \sigma)(\gamma n^*)^{-2} + (\gamma a)^2 d^{-2} (\overline{Z}_m - \sigma)^2 \zeta^{-3}\right\}\right\} \\ &= \sigma + (m-1)(\gamma n^*)^{-1} E(\overline{Z}_m - \sigma) - \gamma a (\gamma n^*)^{-2} (m-1) E(\overline{Z}_m - \sigma)^2 \\ &\quad + (m-1)(\gamma a)^2 E\left\{(\overline{Z}_m - \sigma)^3 \zeta^{-3}\right\} \\ &= \sigma - (\gamma a^2)^{-1} d^2 + (m-1)(\gamma a^2)^2 d^{-4} E\left\{(\overline{Z}_m - \sigma)^3 \zeta^{-3}\right\} \end{aligned}$$

It remains to evaluate the remainder term as follows. We consider two cases:

1. If $(\gamma n^*) \le \zeta \le (M_1 - 1)$,

then have

$$(m-1)(\gamma a^2)^2 d^{-4} E\{(\bar{Z}_m - \sigma)^3 \zeta^{-3}\}$$
$$\leq (m-1)(\gamma a^2)^2 d^{-4} E\,(\bar{Z}_m - \sigma)^3 / (\gamma a^2 d^{-2})^3 = o(d^2)$$

2. If $(M_1 - 1) \leq \zeta \leq (\gamma n^*)$, we write

$$(m-1)(\gamma a^2)^2 d^{-4} E\{(\bar{Z}_m - \sigma)^3 \zeta^{-3}\} \leq (m-1)(\gamma a^2)^2 d^{-4} E\{(\bar{Z}_m - \sigma)^3 (M_1 - 1)^{-3}\}$$
$$\leq (m-1)^{-2} (\gamma a^2)^2 d^{-4} E\{(\bar{Z}_m - \sigma)^3\} = o(d^2) \tag{2.9}$$

Finally, we get

$$E(\hat{\sigma}_{N_1}) = \sigma - d^2 (\gamma a^2)^{-1} + o(d^2);$$

which proves part (i) of Lemma 1. Parts (ii) and (iii) can be established in a similar manner. Taylor expansion, parts (i), (ii), and (iii) prove part (iv) of Lemma 1. Part (v) is direct application of part (iv), thus we omit any further details for brevity.

It is clear from (i) of Lemma 1 that $\hat{\sigma}_{N_1}$ is under estimating σ, the amount of bias depends on the constants a and γ. It is interesting to see any reduction for bias due to the information gained by the third stage. We will elaborate on such improvement, as we develop results concerning the third stage.

The next Theorem provides asymptotic moments of the final stage sample size N.

Theorem 1. *In the context of the rule (2.1)-(2.2), we have, as* $d \to 0$,

$$(i)\, E(N) = n^* - \gamma^{-1} + 1/2 + o(1)$$
$$(ii)\, Var(N) = n^* \gamma^{-1} + o(d^{-2})$$
$$(iii)\, E\,|N - n^*|^3 = o(d^{-2}).$$

Proof of Theorem 1. Following Hall (1981), we write,

$$N = [a^2 \hat{\sigma}_{N_1} / d^2] + 1$$
$$= a^2 \hat{\sigma}_{N_1} / d^2 + \Phi_d$$

where,

$$\Phi_d = 1-(a^2 d^{-2}\hat{\sigma}^2{}_{N_1} - [a^2\,\hat{\sigma}^2 N_1 / d^2])$$

is asymptotically uniform over $(0,1)$ as $d \to 0$ (see Hall (1981, pp 1237) for details). It follows that

$$E(N)=a^2 d^{-2} E(\hat{\sigma}_{N_1})+1/2+o(1).$$

We then apply (iv) of Lemma 1 to obtain (i) of Theorem 1. To prove (ii), we consider

$$Var(N)=Var(a^2 d^{-2}\hat{\sigma}_{N_1}+\Phi_d)$$
$$=Var(a^2 d^{-2}\hat{\sigma}_{N_1})+Var(\Phi_d)+2a^2 d^{-2} Cov(\hat{\sigma}_{N_1},\Phi_d)$$

Using Cauchy–Schwartz inequality, we get

$$\left\{Cov(\hat{\sigma}_{N_1},\Phi_d)\right\}^2 \le Var(\hat{\sigma}_{N_1})\,Var(\Phi_d)=o(d^{-2}),\quad as\quad d\to 0.$$

which shows that the two random variables $\hat{\sigma}_{N_1}$ and Φ_d are asymptotically uncorrelated. Therefore,

$$Var(N)=a^2 d^{-2}\,Var(\hat{\sigma}_{N_1})+o(d^{-2}).$$

Next, we use (i) and (ii) of Lemma 1 to obtain part (ii) of Theorem 1. Part (iii) of Theorem 1 can be established from (iii) of Lemma 1, which completes the proof of Theorem 1.

It would be of interest for further analysis in subsequent section to find second order approximation to a continuous differentiable function of N in the following Theorem 2.

Theorem 2. *In view of the rule (2.1)–(2.2), let g be continuously differentiable function in a neighborhood around n^* such that*

$$\sup_{n\ge m}|g'''(n)|=O(|g'''(n^*)|),$$

then,

$$E(g(N))=g(n^*)+(2\gamma)^{-1}\left\{(\gamma-2)g'(n^*)+n^* g''(n^*)\right\}+o(g'(d^{-2})),\ as\ d\to 0$$

Proof of Theorem 2. We expand $g(N)$ around n^* in Taylor series and take the expectation throughout to obtain

$$E(g(N)) = g(n^*)+E(N-n^*)g'(n^*)+1/2E(N-n^*)^2 g''(n^*)+1/6E\{(N-n^*)^3 g'''(v)\}$$

where v is a random variable lies between N and n^* and primes mean derivatives. We then apply parts (i) and (ii) of Theorem 1 and write

$$E(g(N)) = g(n^*)+(2\gamma)^{-1}\{(\gamma-2)g'(n^*)+n^* g''(n^*)\}+1/2\,E\{(N-n^*)^3 g'''(v).$$

It remains to evaluate the remainder term. Using part (iii) of Theorem 1, it follows that,

$$1/2E\{(N-n^*)^3 g'''(v)\} \le 1/2 \sup_{n\ge m} |g'''(n)|\, E|N-n^*|^3 = O(|g'''(n^*)|)o(d^{-2}) = o(|g'(d^{-2})|).$$

Lemma 2. *In view of the three-stage rule (2.1)-(2.2) we have as* $d \to 0$,

$$(i)E(\hat{\sigma}_N) = \sigma - a^{-2}d^2 + o(d^2)$$
$$(ii)E(\hat{\sigma}_N - \sigma)^2 = n^* + 3/2 + 2\gamma^{-1} + o(1).$$

The proof of Lemma 2 is direct application of Lemma 3 of Almahmeed et. al. (1998, pp 547), thus we omit details. It follows from (i) of Theorem 1 that the third stage indeed reduced the bias noticed in (i) of Lemma 1 above.

The main result of this study is given in the following theorems. To study the asymptotic foregone risk of the triple sampling procedure (2.1)-(2.2), as in Chow and Robbins (1965), we define the asymptotic foregone risk of the triple sampling procedure as

$$\omega = E(L_N(A)) - E(L_{n^*}(A))$$

Theorem 3. *Under the quadratic loss function (1.2), the foregone risk of the procedure (2.1)-(2.2) is given by* $\omega = 2+\gamma^{-1}+o(1)$ *as* $d \to 0$.

Proof of Theorem 3. We first write the foregone risk of the triple sampling procedure as

$$\omega = a^2 d^{-2} E(\hat{\sigma}_N - \sigma)^2 + E(N) - 2n^*$$

Hence, we use (ii) of Theorem 2 and (i) of Theorem 1 to affirm the statement of Theorem 3. It is clear from Theorem 3, that the cost of using the triple sampling procedure (2.1)-(2.2) instead of the fixed sample size procedure is equivalent to $2+\gamma^{-1}$. It is also of interest to affirm that the coverage probability that σ lies within the fixed with confidence interval is at least the nominal value as indicated by the following Theorem.

Theorem 4. *The asymptotic coverage probability of the confidence interval for* σ *that is associated with triple sampling procedure (2.1)−(2.2) is given by*

$$P\{|\hat{\sigma}_N - \sigma| \le d\} = (1-\alpha) - a\,\phi(a)\left(a^2 - 2\gamma + 5\right)(2\gamma n^*)^{-1} + o(d^2) \quad as\ d \to 0$$

where $\phi(\cdot)$ *is the standardized normal p.d.f.*

Proof of theorem 4: The coverage probability of the triple sampling confidence interval for σ can be written as $P\{|\bar{Z}_N - \sigma| \le d\} = \sum_{n=m}^{\infty} P\{|\bar{Z}_N - \sigma| \le d | N = n\} P(N = n)$. Anscombe (1952) central limit theorem also provides

$$\left(\frac{\bar{Z}_N - \sigma}{\sigma}\right)\sqrt{N} \xrightarrow{\ell} N(0,1) \text{ as } d \to 0$$

it follows that

$$P\{|\bar{Z}_N - \sigma| \le d\} = E\left(2\Phi\left(\frac{d\sqrt{N}}{\sigma}\right) - 1\right)$$

where $\Phi(\cdot)$ is the standardized cumulative distribution function of the normal distribution. Now, apply theorem 2 above, then the statement of theorem 4 is immediate.

The quantity $\left(a^2 - 2\gamma + 5\right)/(2\gamma)$ appears in the statement of theorem 4 is known as the cost of not knowing σ (the cost of ignorance).

Since the results given in section 2 in this paper are asymptotic in nature, it is vital to assess the small to moderate sample size performance of the above procedure in real life applications.

3. Moderate Sample Size Performance

A series of 5000 Monte−Carlo simulations were generated to present a sense regarding the , small to moderate sample size behavior of the integrated three−stage sampling scheme (2.1)−(2.2). As has been mentioned in Section 1, the procedure provides a fixed width confidence interval for σ with at least $100(1-\alpha)\%$ confidence coefficient and at the same time provides a point estimate for σ under squared error loss function with linear sampling cost. In the simulations, we allowed the optimal sample size to vary from small $n^* = (5,10,20,25,30,50,75,100,150,200,500,1000)$ to large. We fixed $\theta = 1$ and $\sigma = 1$ in all simulations. The design factor γ assumed values $0.3, 0.5, 0.7$, and 0.8 to study the effect of changing this factor on the final stage sample size. The starting sample size m took values $5,10,15$ and 20 to study the effect of changing the initial sample size on the quality of the three−stage estimates. The

confidence coefficient was permitted to have the nominal values 0.90, 0.95 and 0.99. The following tables present the simulation results for selected combinations of γ, m, α, and n^*.

In each case, we start the procedure by generating m samples from Pareto distribution with $\theta = 1$ and $\sigma = 1$. Hence, estimates for $\sigma\,(\hat{\sigma}_m)$ and $\theta\,(X_{m(1)})$ are computed to initiate the three stage rule in (2.1)−(2.2). Of course, these estimates are updated during the course of estimation. Upon observing $N(X_{N(1)})$ and $\hat{\sigma}_N$ at the final stage we store these values for further use. The process is repeated 5000 times. Finally estimates $\overline{N}$ of n^*, $\hat{\sigma}_N$ of σ, $\hat{\varpi}$ for ϖ, and $\hat{p}$ for $(1-\alpha)$ are computed based on 5000 simulations.

Close investigation to the simulation results presented in Tables 1−5 illustrates several facts regarding the performance of the triple sampling procedure (2.1)−(2.2). Generally speaking, it is noticed that in all tables the triple sampling estimates of σ is very close from the targeted value 1. Its standard error decreases with the increase of n^*. Also, the targeted coverage probability attains the nominal value. Two important remarks regarding the role played by the size of the pilot sample m and the design factor γ. We have notice that if the starting sample size happens to be chosen smaller than the optimal, this will lead to a substantial over sampling (compare $\overline{N}$ and n^*) presented in the tables. The amount of bias decreases as n^* shifts away from the value of m. A slight savings in the triple sampling sample size is notice for larger values of n^*. The simulations also reveals that if γ is less than 0.5 a substantial over sampling occurs especially with smaller values of n^*. But if γ is greater than 0.5 a substantial over sampling occurs with larger values of n^*. Therefore, to avoid the problem of under or over sampling, one should guess a value of m not very much smaller nor very much larger than n^* "compare the amount of bias in tables 2 and 4". Many authors including Hall (1981), Hamdy (1988), Hamdy et al. (1988), Hamdy et al. (1989), and Hamdy and Pallotta (1987) argued that the choice of m to be 5, 10, 15, or 20 is reasonable for practical implementation. These findings support the assumption that $Sup\frac{m}{n^*} < \gamma$ set forward by Hall (1981). The choice of design factor γ to be 0.5 is reasonable for practical use "compare $\left|\overline{N} - n^*\right|$ in table 2 and 5". It is worth mentioning that the asymptotic foregone risk is always negative, which illustrates that the triple sampling procedure performs better than the fixed sample size procedure has σ been known. This could be due to the dependence between the three−stage sample size N and the parameter estimate $\hat{\sigma}_N$.

Acknowledgement

We would like to express our gratitude to Mr. Yousef, A. S. for conducting the simulation study of the three−stage procedure presented in this work.

Table (1)
$m = 5,\ \gamma = 0.3,\ \alpha = 0.05$

n^*	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\sigma}_N$	$S.E.(\hat{\sigma}_N)$	$\hat{\omega}$	$\hat{P}$	$\lvert\overline{N} - n^*\rvert$
5	7.25	0.050	1.0047	0.3367	−2.75	0.9978	2.25
10	10.90	0.089	0.9983	0.3247	−9.10	0.9480	0.90
15	14.57	0.128	1.0034	0.3188	−15.43	0.8812	0.43
20	18.26	0.166	1.0016	0.2973	−21.74	0.8364	1.74
25	22.10	0.203	0.9954	0.2782	−27.90	0.8460	2.90
30	26.56	0.239	0.9993	0.2678	−33.44	0.8420	3.44
50	43.34	0.361	0.9971	0.2266	−56.66	0.8348	6.76
75	66.51	0.505	1.0004	0.1909	−83.49	0.8464	8.49
100	92.44	0.629	1.0027	0.1576	−107.56	0.8876	7.56
200	189.76	0.996	0.9999	0.1022	−210.24	0.8992	10.24
500	493.04	1.822	0.9999	0.0550	−506.96	0.9364	6.96
1000	1000.92	2.886	0.9999	0.0341	−999.08	0.9404	0.92

Table (2)
$m = 10,\ \gamma = 0.5,\ \alpha = 0.05$

n^*	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\sigma}_N$	$S.E.(\hat{\sigma}_N)$	$\hat{\omega}$	$\hat{P}$	$\lvert\overline{N} - n^*\rvert$
5	10.33	0.018	0.9921	0.2830	0.33	0.9982	5.33
10	12.76	0.059	1.0012	0.2503	−7.24	0.9912	2.76
15	15.98	0.093	0.9981	0.2505	−14.02	0.9620	0.98
20	19.93	0.129	1.0054	0.2450	−20.07	0.9316	0.07
25	23.96	0.160	1.0002	0.2278	−26.04	0.9024	1.04
30	28.20	0.194	0.9990	0.2259	−31.80	0.8720	1.80
50	46.97	0.301	1.0003	0.1871	−53.03	0.8814	3.03
75	70.86	0.398	0.9974	0.1527	−79.14	0.8912	4.14
100	96.28	0.483	0.9967	0.1278	−103.71	0.9020	3.72
150	147.85	0.657	1.0006	0.0991	−152.16	0.9192	2.15
200	199.66	0.756	1.0017	0.0800	−200.34	0.9320	0.34
500	507.41	1.559	0.9997	0.0462	−492.59	0.9418	7.41
1000	1028.43	2.703	1.0001	0.0328	−971.57	0.9426	28.41

Table (3)
$m = 15$, $\gamma = 0.5$, $\alpha = 0.05$

n^*	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\sigma}_N$	$S.E.(\hat{\sigma}_N)$	$\hat{\omega}$	$\hat{P}$	$\lvert\overline{N} - n^*\rvert$
5	15.03	0.005	0.9995	0.2568	5.03	0.9996	10.03
10	15.98	0.038	1.0065	0.2221	−4.02	0.9986	5.98
15	18.52	0.076	1.0002	0.2121	−11.47	0.9868	3.52
20	21.73	0.105	0.9952	0.2069	−18.27	0.9682	1.73
25	25.63	0.136	0.9987	0.2073	−24.38	0.9476	0.63
30	29.53	0.164	1.0009	0.2022	−30.47	0.9232	0.47
50	47.30	0.263	1.0037	0.1722	−52.70	0.8994	2.70
75	71.00	0.359	1.0009	0.1449	−79.00	0.9008	4.00
100	95.67	0.440	1.0021	0.1234	−104.33	0.9068	4.43
150	145.51	0.558	0.9989	0.0955	−154.49	0.9180	4.49
200	197.05	0.654	1.0016	0.0776	−202.95	0.9302	2.95
500	502.77	1.177	1.0007	0.0459	−497.23	0.9412	2.77
1000	1006.64	1.765	0.9998	0.0321	−993.36	0.9458	6.64

Table (4)
$m = 20$, $\gamma = 0.5$, $\alpha = 0.05$

n^*	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\sigma}_N$	$S.E.(\hat{\sigma}_N)$	$\hat{\omega}$	$\hat{P}$	$\lvert\overline{N} - n^*\rvert$
5	20.00	0.002	0.9984	0.2225	10.00	0.9984	15.00
10	20.18	0.016	0.9963	0.2126	0.19	0.9990	10.18
15	21.46	0.052	0.9985	0.1954	−8.54	0.9928	6.48
20	24.15	0.089	1.0007	0.1884	−15.85	0.9830	4.15
25	27.46	0.119	0.9985	0.1856	−22.54	0.9666	2.46
30	31.22	0.147	0.9975	0.1857	−28.78	0.9524	1.22
50	47.33	0.239	0.9966	0.1638	−52.67	0.9034	2.67
75	70.81	0.340	0.9975	0.1327	−79.19	0.9154	4.19
100	96.05	0.418	1.0020	0.1183	−103.95	0.9146	3.95
150	145.27	0.536	1.0002	0.0927	−154.73	0.9230	4.73
200	195.57	0.635	1.0015	0.0790	−204.43	0.9268	4.43
500	497.89	1.056	1.0000	0.0457	−502.11	0.9452	2.11
1000	999.40	1.563	0.9995	0.0319	−1000.59	0.9488	0.60

Table (5)
$m=10,\ \gamma=0.7,\ \alpha=0.05$

n^*	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\sigma}_N$	$S.E.(\hat{\sigma}_N)$	$\hat{\omega}$	$\hat{P}$	$\lvert\overline{N}-n^*\rvert$
5	10.37	0.019	1.0053	0.2906	0.37	0.9982	5.37
10	12.45	0.056	0.9993	0.2490	−7.55	0.9930	2.45
15	16.17	0.101	1.0021	0.2445	−13.83	0.9672	1.17
20	20.49	0.142	1.0027	0.2409	−19.51	0.9352	0.49
25	25.56	0.183	0.9998	0.2298	−24.44	0.9014	0.56
30	30.67	0.216	1.0041	0.2205	−29.33	0.8816	0.67
50	51.22	0.333	0.9992	0.1847	−48.78	0.8844	1.22
75	78.45	0.452	1.0000	0.1504	−71.55	0.9014	3.49
100	105.94	0.604	0.9989	0.1241	−94.06	0.9154	5.94
150	161.22	0.807	0.9991	0.0917	−138.78	0.9288	11.22
200	218.76	1.062	1.0003	0.0748	−181.24	0.9408	18.76
500	550.84	2.300	1.0005	0.0443	−449.16	0.9486	50.84
1000	1108.65	4.710	0.9989	0.0308	−891.35	0.9490	108.65

References

[1] Ahsanullah, M. (1991). On record values from the generalized Pareto distribution distribution, *Pakistan Journal of Statistics*. A, 7, 129−136.

[2] Ahsanullah, M. and Houchens R. L. (1989). A note on record values from a Pareto distribution, *Pakistan Journal of Statistics*. A, ,5, 15−57.

[3] Almahmeed, M. A., Al−Hussainan, A., Son, M. S., and Hamdy, H. I.(1998). Three stage estimation for the mean of a one−parameter exponential family. *The Korean Communications in Statistics*, 5, 539−557.

[4] Almahmeed, M. A., Hamdy,H.I. and Alzalzalah, Y.H. (2002). An Integrated Sequential Inference Approach for the Normal Mean, *Journal of Korean Statistical Society*, 31: 415−431.

[5] Anscombe, F.J.(1953), Sequential estimation. *Journal of Royal. Statistical. Society.*, B, 15, 1−21.

[6] Arnold, B.C. and Press, S.J.(1983). Bayesian inference for Pareto populations, *Journal of Econometrics*, 21, 287−306.

[7] Balakrishnan N. and Ahsanullah M. (1994b). Recurrence relations for single and product moments of record values from generalized Pareto distribution, *Communication in Statistics, Theory and Methods*, 23, 2841−2852.

[8] Chow, Y.S. and Robbins, H.(1965). On the asymptotic theory of fixed width sequential confidence intervals for the mea,. *Annals of Mathematical Statistics*, 36, 457−462.

[9] Hall, P.(1981). Asymptotic theory of triple sampling for sequential estimation of a mean, *Annals of Statistics*, 9, 1229−1238.

[10] Hamdy, H.I.(1988). Remarks on the asymptotic theory of triple stage estimation of the normal mean, *Scandinavian Journal. of Statistics*, 15, 303−310.

[11] Hamdy, H.I., Mukhopadhyay, N., Costanza, M.C. and Son, M.S.(1988). Triple stage point estimation of the exponential location parameter, *Annals of the Institute Statistical Mathematics*, 40, 785−797.

[12] Hamdy, H.I., Son, M.S., Al−Mahmeed, M. and Nigm, A.(1989). Three−stage point estimation procedure for the exponential location parameters, *Metron*, XLVII, 279−294.

[13] Hamdy, H.I. and Pallotta, W.J.(1987). Triple sampling procedure for estimating the scale parameter of Pareto distribution, *Communication in Statistics, Theory and Methods*, 16, 2155−2164.

[14] Liang T. C. (1993). Conversion rates of empirical Bayes estimation of the scale parameter in Pareto distribution, *Computational Statistics and data analysis*, 35-45.

[15] Lombard, F. and Swanepoel, J.W.H.(1978). On finite and infinite confidence sequences, *South African Journal of Statistics*, 12, 1-24.

[16] Malik, H.J.(1970). Bayesian estimation of the Paretian index, *Scandinavian Aktuarietidskr*, 53, 6-9.

[17] Mann, N.R., Schafer, R.E. and Singpurwalla, N.D.(1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York, John Wiley and Sons.

[18] Mukhopadhyay, N.(1985). A note on three-stage and sequential point estimation procedures for a normal mean, *Sequential analysis*, 4, 311-319.

[19] Mukhopadhyay, N. and Eko, M.E.(1985). Sequential estimation problem for the scale parameter of a Pareto distribution, *Technical Report, Oklahoma State University*.

[20] Mukhopadhyay, N. and Hamdy, H.I., Al-Mahmeed, M. and Costanza, M.C.(1987). Three-stage point estimation for a normal mean, *Sequential Analysis*, 6, 21-36.

[21] Muniruzzaman, A.N.M.(1986). Estimation of the parameters of Pareto distribution. A Bayesian approach. Statistics for development proceedings, *Institute of Statistical Research and Training University of Dacca*.

[22] Nigm, A. and Hamdy, H.I.(1987). Bayesian prediction bounds for the Pareto lifetime model, *Communication in Statistics, Theory and Methods.*, 16, 1761-1772.

[23] Pareto, V.(1897). Cours d'Economic Politique, *Lausanne and Paris*: Rouge and Cie.

[24] Sinha, S.K. and Howlader, H.A.(1980). On the sampling distributions of Bayesian estimators of Pareto parameters with proper and improper priors and associated goodness of fit, *Technical report #103, Department of Statistics, University of Manitoba, Canda*.

[25] Swanepoel, J.W. and van Wyk(1982). Fixed width confidence intervals for the location parameter of an exponential distribution, *Communication in Statistics, A*, 11, 1279-1289.

[26] Stein, C.(1945). A two stage sample test for a linear hypothesis whose power is independent of the variance, *Annals of Mathematical Statistics*, 16, 243-252.

[27] Wang, Y.H.(1973). Sequential estimation of the scale parameter of the Pareto distribution, *Communication in Statistics, Theory and Methods*, 2, 145-154.

[Received October 2003, Accepted November 2003]